



ANALYSER DES DONNÉES

Nadine Mandran – Juin 2022

Laboratoire Informatique de Grenoble



LIVE ANALYSIS

- Si l'analyse de données était un animal, elle serait

L'ANALYSE ET VOUS

- Quelles difficultés rencontrées vous ?
 - En quali ?
 - En quanti ?

ANALYSER

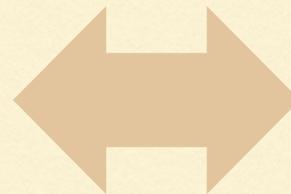
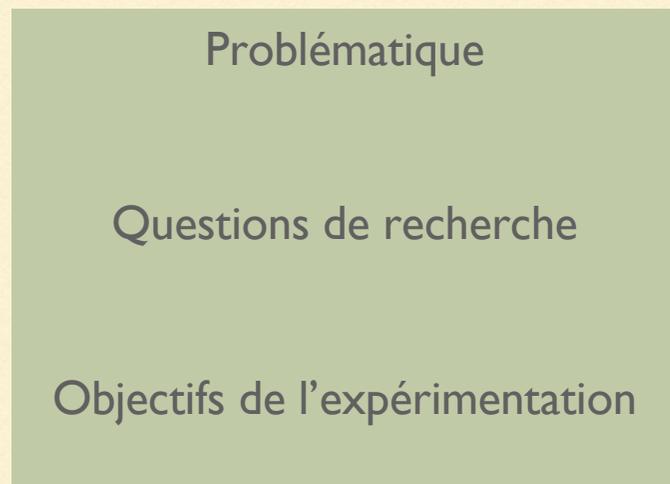
- L'analyse des données est la fin d'un cycle
- Problématique, Question de recherche, Contribution, Protocole expérimental, Production des données, ANALYSE, Conclusion
- Une bonne analyse c'est avant tout de bien poser le problème et de produire correctement des données ...

PROTOCOLE EXPERIMENTAL

- Recenser l'ensemble des étapes qui vont être conduites pour construire et évaluer la proposition de recherche avec des utilisateurs
- Garantir la traçabilité des différentes étapes de l'étude menée

http://thedre.imag.fr/wp-content/uploads/2022/02/V5_Rediger_protocole_experimental.docx

ANALYSE DES DONNEES QUALITATIVES



DONNÉES ?

- une **donnée Existante** hors du contexte de la recherche que le chercheur peut mobiliser pour répondre à ses questions. (e.g. un support de cours)
- Une **donnée Mesurée** données construites par et pour le chercheur sur la base de travaux théoriques dans une démarche scientifique (e.g. une note de d'utilisabilité).

DONNÉES ?

- **Factuelles** résultent de l'observation de faits ou d'activités. Elles sont mesurées par des traces d'activité sur la plateforme ou des captures vidéo de personnes en situation.
- **Déclaratives** reposent sur l'expression des personnes sur leurs expériences et leurs perceptions. Elles sont recueillies par des questions dans des questionnaires ou lors d'entretiens.

DONNÉES ?

- **Qualitatives** : Peu de connaissances, Comprendre, identifier des phénomènes, Etudier la diversité - Nombre de personnes faibles mais divers profils.
- **Quantitatives** : Connaissances sur le phénomène, Dénombrer, quantifier, tester - Méthode d'échantillonnage : quota, aléatoire, plan d'expérience

QUALITATIVES

Objectifs

- Accéder au sens (Duyck, 2003) et/ou aux représentations.
- Comprendre
- Identifier des phénomènes, des notions inconnues
- Appréhender des notions inconnues
- Connaître en profondeur et avec précision un phénomène
- Observer la diversité

« Echantillon »

- Des profils divers pour augmenter la diversité
 - Par de représentativité de la diversité
-

QUALITATIVES

Les méthodes de PRODUCTION de données QUALITATIVES

- Entretien semi- directif
- Focus-group ou groupe de discussion
- Auto-confrontation
- Journal de bord et « Social probes »
- Observation

Les données produites

- Audio
 - Vidéo
 - Carnets de notes
 - Schéma...
-

ANALYSE DES DONNEES QUANTITATIVES

Comment faites vous ?

SYNTHÈSE A CHAUD

Une synthèse même rapidement écrite après un entretien ou une séance de focus group est essentielle, elle permet de gagner du temps lors de l'analyse des données.

RESUME DES ENTRETIENS

ETUDE : NOM DE L'ETUDE

Enquêteur(rice) : Nom, Prénom

Enquêté(e) : Nom, Prénom

Date et heure de rdv :

Nom Mnémonique :



Ton de l'entretien	
Evénements survenus	
Résumé :	
Reprendre les thèmes de la grille d'entretien	
Noter ce que le sujet évoque pour chacun des thèmes	
Eléments en lien avec l'objectif de l'expérimentation et de la thèse	
Eléments novateurs ou marquants :	
Note de 0 à 5 0 peu intéressant, 5 très intéressant	

ANALYSE QUALITATIVE

Deux familles d'analyse des données

- **Analyse thématique : manuelle ou semi-automatisée**
 - Construire une liste de thème en lien avec l'objet de l'étude
 - Relever les extraits de phrases qui évoquent le thème (manuelle ou semi-automatisée : Taguette, Elan, Atlas TI, ..)
 - **Analyse du contenu, du discours, automatisée**
 - Analyses statistiques du corpus sur la base d'étude du champ lexical (Sphinx), sur la base de la syntaxe ou de la grammaire (Alceste, Rtemis, Rainette)
-

ANALYSE THÉMATIQUE

P.Paille et A.Muchielli, L'analyse qualitative en sciences humaines et sociales, 2012.

ANALYSE THÉMATIQUE

Ce n'est

Ni un codage des données => trop réducteur difficile de faire une synthèse

Ni une quantification des thèmes => pas d'échantillonnage statistique, pas produit pour cela

ANALYSE QUALITATIVE

Retranscrire les entretiens (1h => 6h) , des logiciels de retranscriptions :
NOOTA <https://noota.io/>

A partir de la problématique et de la grille d'entretien, créer une grille thématique,

A partir de la **retranscription des entretiens**, identifier les thèmes qui correspondent à la problématique et aussi les thèmes qui émergent

Les thèmes sont larges, il convient de créer des sous-thèmes

Chacun des sous-thèmes est illustré par des verbatims (extraits du discours associé au code du sujet ayant prononcé de discours)

Synthèse pour chaque thème

Synthèse globale

ANALYSE THÉMATIQUE

Comment faire ?

1- A partir des questions de recherche, des questions du guide d'entretien, élaborer une grille d'analyse

Pour vous aider utilisez ces trois niveaux d'analyse :

- **Descriptif** : « *il a pris son manuel scolaire* »
- **Interprétatif** : « *là je crois qu'il a compris* »
- **Explicatif** : « *Cet élève a vraiment progresser car il a beaucoup travailler la géométrie ce trimestre* »

Pour les hypothétiques-déductifs, pour vous aider utilisez des modèles d'analyse existants

ANALYSE THÉMATIQUE

Comment faire ?

2- A la lecture des retranscriptions ou à l'écoute des enregistrements, Identifier des thèmes pour répondre à vos questions, à affirmer ou à rejeter vos hypothèses .

3- Identifier aussi des nouveaux thèmes auxquels vous n'auriez pas pensés (raisonnement abductif)

4- Garder les *verbatim* (extraits de phrases) pour « prouver » ce que vous avez identifié et le code de la personne associée

5- Elaborer ensuite une synthèse pour chaque thème

6- Mettre en lien avec vos questions sou hypothèses

ANALYSE QUALITATIVE

Exemple : Une étude sur l'impact des smartphones et des tablettes

Thème « Utilisation des smartphones »

Sous –thème : « utilisation générale »
« *je fais tout avec mon téléphone* » (S4)

Sous thème « type d'application »

1- utilisation pour télécharger des films
2- utilisation pour voir des photos avec des amis »
« *c'est pour voir les photos que l'on en fait entre potes* (S6) »

Thème nouveau de « remplacement des tablettes par les téléphones »
« *je ne l'utilise plus elle est tombée en panne , j'ai pas changé mon téléphone st
suffisant* » (S4)

ANALYSE QUALITATIVE

Points forts et Points faibles

Outil : Grille qui liste les éléments relatifs à une proposition et qui à deux colonnes une pour noter les points forts et l'autre les points faibles

Remarque : Outil de production pour une situation expérimentale où le nombre de personnes est réduit.

Exemple :

Table Number : First and last name :

<i>List the strengths and weaknesses on the following</i>	Strengths	Weaknesses
the work performed with the method	+	-
the collaborative work enabled with the method	+	-
The creativity enabled with the method	+	-
the serious gaming of the method	+	-
The software compared to card game	+	-

ANALYSE QUALITATIVE

Points forts et Points faibles

Méthode :

Saisir l'ensemble des réponses
: Ligne=participants, Colonnes
= réponse aux questions, 2
colonnes par éléments

Experimentation menée en 2011 lors de l'école d'été sur les GBL. LIG/METAH Pernin et all							
saisie	suji	tab	ses	The collaborative work enabled with the method		The method ability to give new ideas	
				What you like +	What you don't like -	What you like +	What you don't like -
CM	1	1	2	It's easy to work collaboratively	Difficult to agree in the selected cards	very intuitive to create many different ideas	
CM	2	1	2	immediate reviewing	maybe it can be difficult to make the game with people with different backgrounds	the fact of having one expert facing newbies on the subject	
CM	3	1	2	The cards are good support to communicate with other members of the team. Rapid modification / iteration	If you move the table everything moves and you have to replace the cards / jewels	the existing game principles and cards help us to set up some foundations	can also avoid some possible activities or scenarios because of the formalism. If / esle ? Random ?
CM	1	2	2	great for triggering discussion. The terminology might need to be tightened. Capitalize might not work that well.	Potential for one person to dominate the interaction physically (but not completely)	it was very helpful. I have a lot of new ideas I may follow up. Perhaps more creative game scenarios in activity 1 would be useful for inspiration.	need more time afterwards for reflection + exploring ideas that arise
CM	2	2	2	we shared many different ideas	we should take some roles so we could be more organized	we produced a lot of ideas	
CM	3	2	2	cards facilitate greatly the collaborative approach.		this method brings new ideas because students start on groupware gaming activities and their synchronisation and communication issues.	the method should be extendible ; introduce profiles.
CM	4	2	2	I like combination of different opinions of participants		i like to see differents point of view	
CM	1	2	3	échanges nombreux ; matériel (forte flexibilité dans la manipulation des différents éléments) a aidé à la collaboration	auto-régulation => en cas de désaccord, pas de moyen prévu pour aider à trouver des compromis	grandes variétés de choix; Prévoit différentes méthodes de travail, etc.	cadre tout de même très normatif trop contraignant, il est difficile de se conformer aux cartes dans la phase de conception globale
CM	2	2	3	la répartition des rôles ; le retour collectif	pas assez détaillé dans la répartition des rôles	l'encadrement proposé	
CM	3	2	3	visualisation de la progression, des itérations	n'aide pas forcément à la collaboration quand les mécaniques de jeu sont machées dès la début	on essaie de partir des ressorts de jeu	donner des exemples de jeu pour faciliter les idées

ANALYSE QUALITATIVE

Points forts et Points faibles

Méthode :

Faire une « analyse thématique simplifiée » et une synthèse

The collaborative work enabled with the method

Les points positifs

3 cases vides (dont 2 vides aussi côté points négatifs)

1. Concernant la méthode		
La méthode facilite et stimule la collaboration des formateurs	great for triggering discussion The method promotes collaboration. Everyone participating, Good to develop team spirit Easy, nice, good way	14
Elle permet de tous participer, de partager des idées différentes et d'être créatifs collectivement	very good- helped to see each participants point of view and thus come up with a collective work Effective debating, compromising, selection and implementation We all participate in the creation process by bringing ideas and say what we think is not good at other ideas emergence of the solution by speaking and interacting confront, exchange, share ideas, créativité, débat, opinion, brainstorm, open discussion friendly	17

ANALYSE QUALITATIVE

Schéma

Créer une grille d'analyse pour savoir ce qui doit être repéré par rapport à la problématique

Photographier chacune des productions

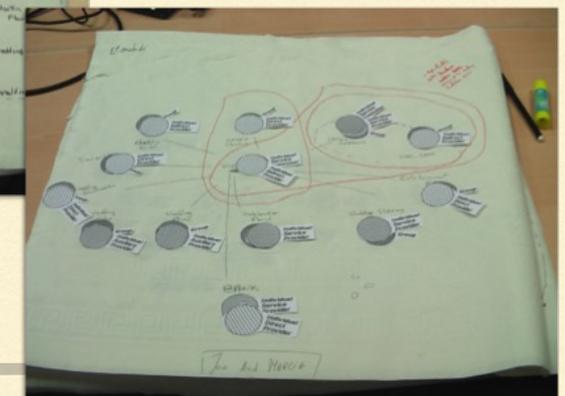
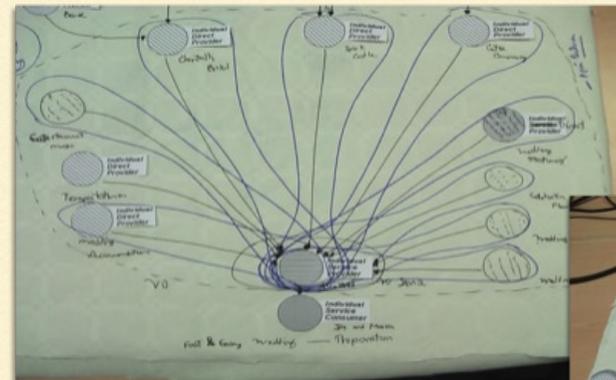
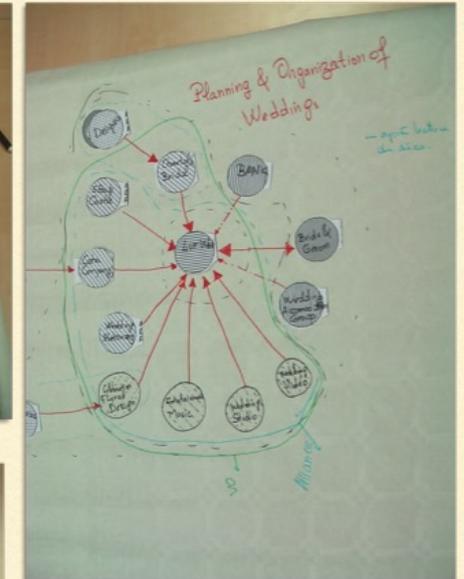
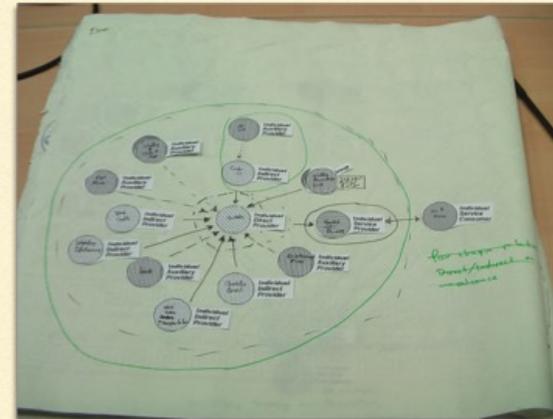
Repérer les ressemblances/les différences entre les représentations

Repérer ce qui confirme les objectifs de l'étude

Repérer les apports de la représentation des sujets par rapport à la proposition

Visionner et écouter les films

Conduire une « analyse thématique simplifiée »



ANALYSE QUALITATIVE

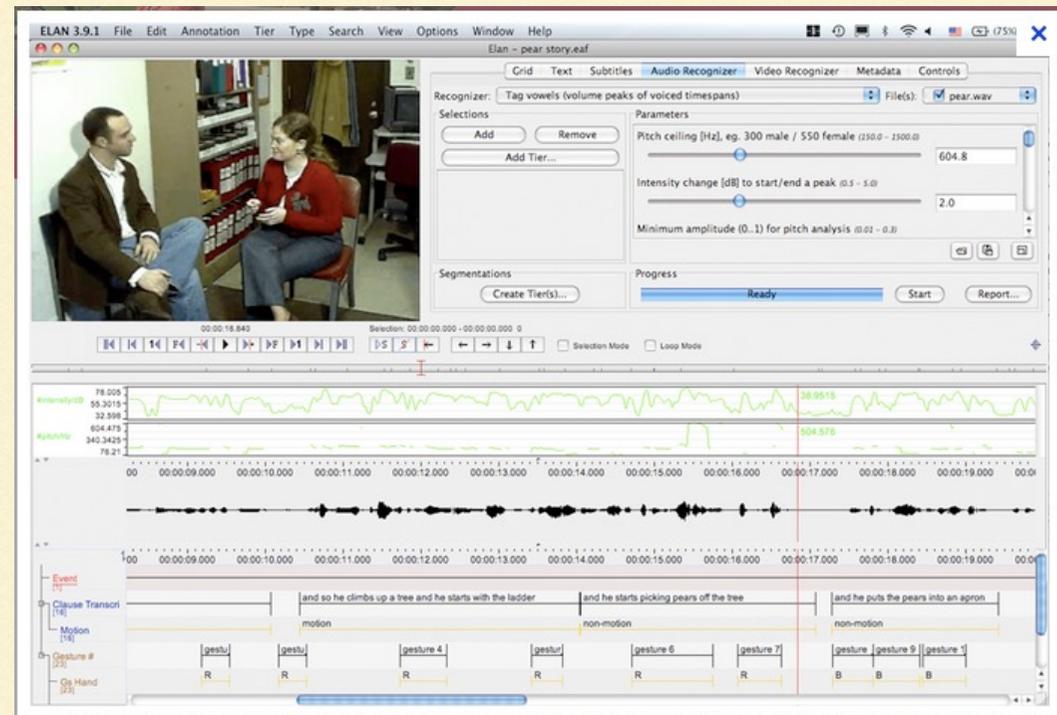
Video

Créer une grille d'analyse pour savoir ce qui doit être repéré par rapport à la problématique.

Utiliser des logiciels d'annotations de video

Analyser les annotations
Statistiques descriptives
Extraction de verbatims

Faire une synthèse



QUANTITATIVES

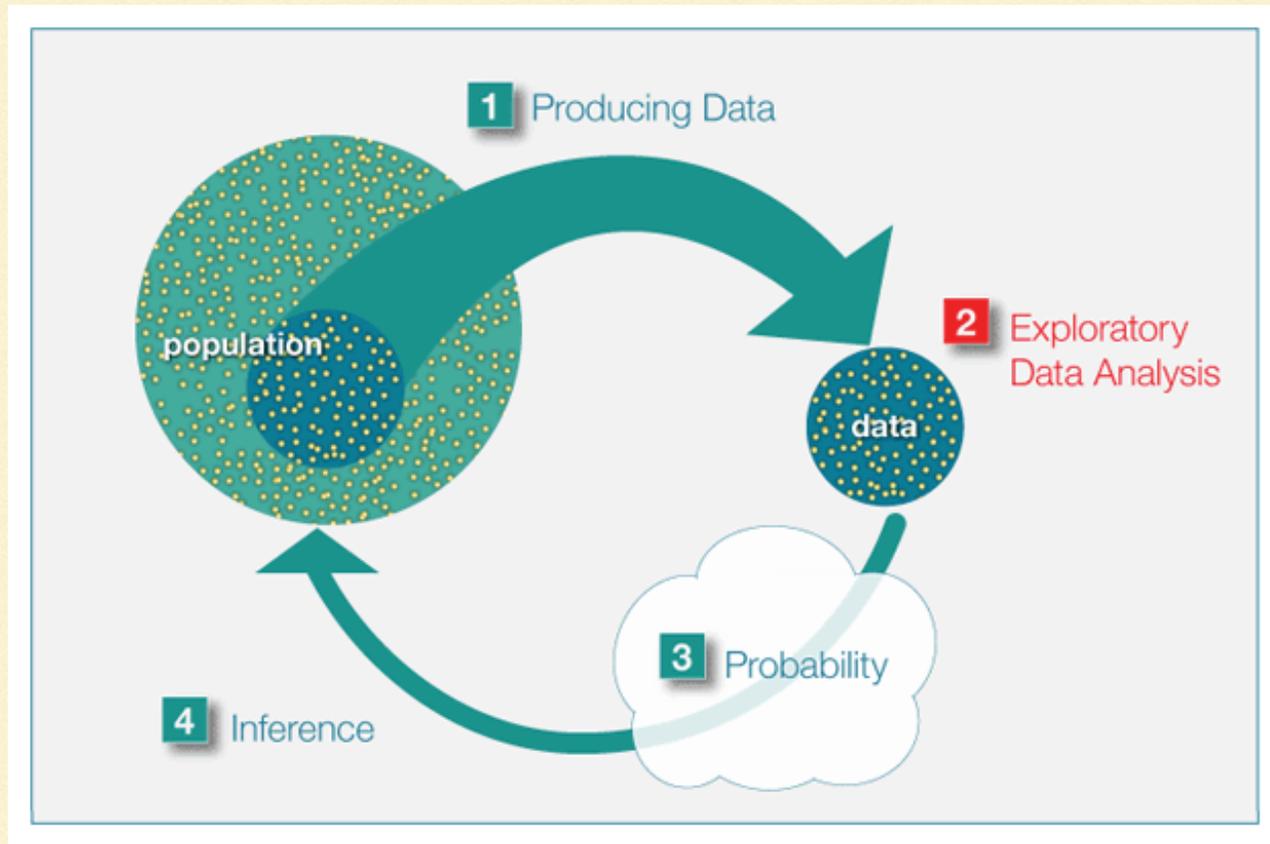
Objectif

- Quantifier
- Evaluer, valider
- Modéliser, prévoir
- Diagnostiquer

Echantillonnage

- Méthode d'échantillonnage aléatoire
 - Méthode échantillonnage par quotas
 - Plans d'expériences : groupe contrôle vs groupe expérimental (Post-Positivisme)
 - Recensement
-

QUANTITATIVES



Source : <https://oli.web.cmu.edu/openlearning/forstudents/freecourses/statistics>

ECHANTILLONNAGE, INDIVIDUS

- **Sans description de la méthode d'échantillonnage, les résultats ne sont valides que pour les données recueillies mais la généralisation n'est pas possible.**

Les quotas

- Caler les résultats sur les données du recensement de la population ou sur des informations globales sur la population étudiée.
- Le plus pratiqué car moins coûteux

La méthode aléatoire

- Tirer au hasard dans une base de sondage.
- La probabilité d'inclusion d'un individu est connue sans biais.
- Coûteux car l'individu tiré au hasard ne doit pas être abandonné

L'exhaustivité (le recensement)

- Avoir la totalité des individus ou des actions faites par les individus

Le plan expérimental

- Le nombre de sujets est fixé a priori en fonction des facteurs expérimentaux
- Définir la structure de l'expérience

QUANTITATIVES

Les méthodes de PRODUCTION des données quantitatives

- **Questionnaire**

Auprès d'un échantillon correctement échantillonné (en sociologie minimum 500 personnes)

Peut-être administré en face-à-face ou en ligne

- **Capture de traces d'activités**

Avoir une application qui permet de tracer les interactions des utilisateurs

- **Tests utilisateurs**

Avoir une application sur laquelle on va faire des tests précis

Souvent utilisés en IHM, mesure de performance, de précision, erreurs, ...

Les données quantitatives

- **Fichier : tableau ligne*colonne, type CSV , excel, data set de R, etc..**
-

QUANTITATIVES

- VOCABULAIRE DES DONNEES => CHOIX DES METHDOES D'ANALYSE DES DONNEES
 - Ligne : **Individus statistiques**
 - Colonne : **Variables**
 - A découvrir par vous mêmes ...

QUANTITATIVES

- Fichier de données
 - Utilisation d'une plateforme pour former à la construction de protocoles expérimentaux en chimie
 - Impact des niveaux de rétroaction de la plateforme sur l'apprentissage

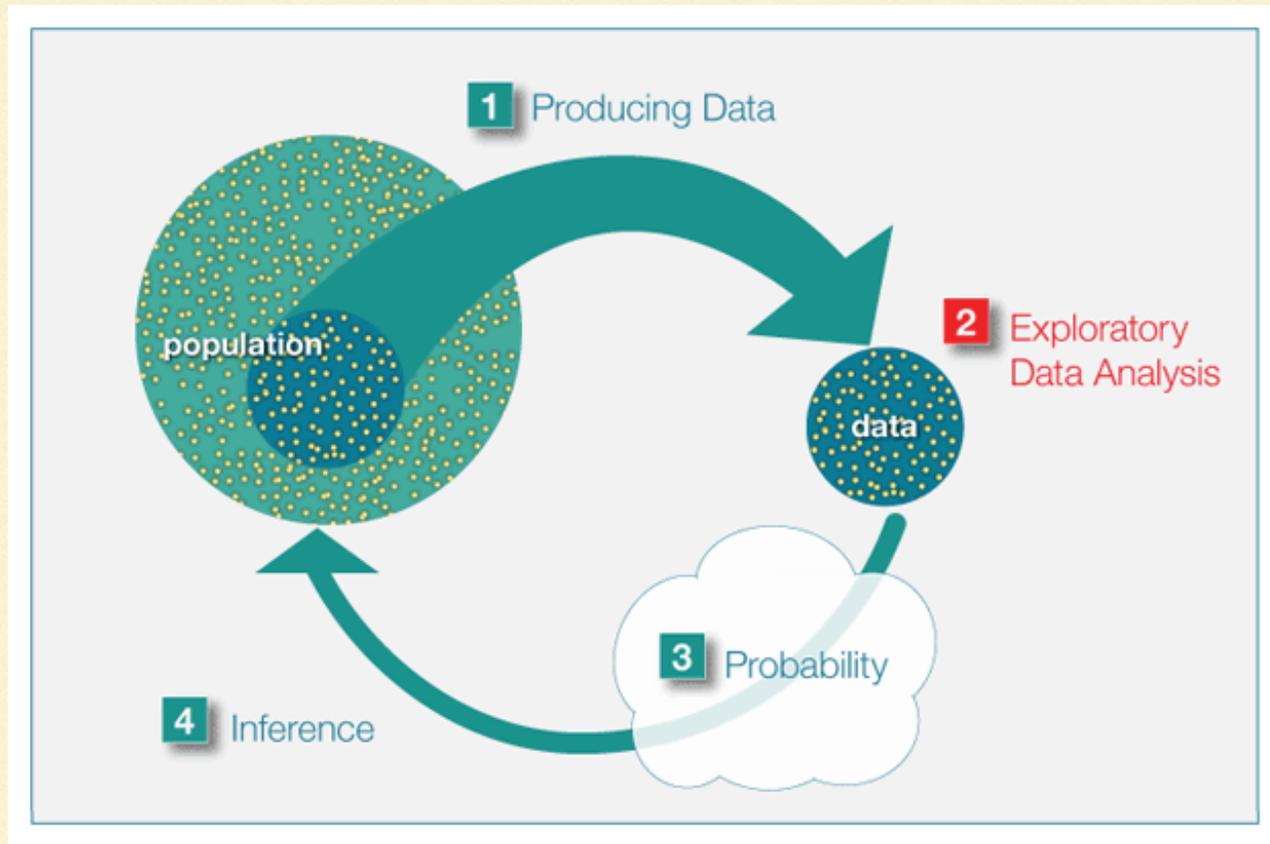
QUANTITATIVES

- . Quel est le **thème**, la sémantique des variables ?
- . **Quelles caractéristiques/quels types** ont les valeurs prises par ces variables ?
- . **Quand** sont collectées les données ?
- . **Qui** les a produites ?
- . **Quel est leur rôle** dans l'expérimentation et l'analyse des données ?
- . Quels sont les **objectifs de recherche** qui pourraient être à l'origine de la création de ces données ?
- . Quelle est la **méthode d'échantillonnage** ?

QUANTITATIVES

- Corrections

QUANTITATIVES



Source : <https://oli.web.cmu.edu/openlearning/forstudents/freecourses/statistics>

ECHANTILLONNAGE, INDIVIDUS

- **Sans description de la méthode d'échantillonnage, les résultats ne sont valides que pour les données recueillies mais la généralisation n'est pas possible.**

· **Les quotas**

- Caler les résultats sur les données du recensement de la population ou sur des informations globales sur la population étudiée.
- Le plus pratiqué car moins coûteux

· **La méthode aléatoire**

- Tirer au hasard dans une base de sondage.
- La probabilité d'inclusion d'un individu est connue sans biais.
- Coûteux car l'individu tiré au hasard ne doit pas être abandonné

· **L'exhaustivité (le recensement)**

- Avoir la totalité des individus ou des actions faites par les individus

· **Le plan expérimental**

- Le nombre de sujets est fixé a priori en fonction des facteurs expérimentaux
- Définir la structure de l'expérience

3 Moments dans la production des données

AVANT

PENDANT

APRES

l'expérience avec une plateforme

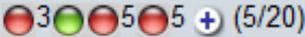
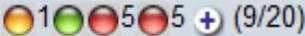
Socio-démographiques
Pré-test
Facteurs expérimentaux
...

La date et heure de l'action
Les actions avec la
plateforme
....

Le post-test
La note d'utilisabilité
La durée de connexion
Le nombre d'actions
Le score (post-test - pre test)
...

QUANTITATIVES

Traces brutes

30	18/01/08	14:50:32	ajout_action	homogeneisation(19) - materiel = ; solution = sol4 ; commentaire = ; étape = 2	
31	18/01/08	14:50:45	modif_ordre		
32	18/01/08	14:50:45	ajout_action	homogeneisation(20) - materiel = ; solution = sol5 ; commentaire = ; étape = 2	
33	18/01/08	14:51:14	modif_ordre		
34	18/01/08	14:51:15	demande_eval	3 61 411	
35	18/01/08	14:52:04	modif_ordre		
36	18/01/08	14:52:04	suppr_produit	sirop de grenadine	
37	18/01/08	14:52:08	demande_eval	2 3 61 411	
38	18/01/08	14:52:55	consult_cours	dosages - onglet_cours	
39	18/01/08	14:54:12	demande_eval	2 3 61 411	
40	18/01/08	14:54:23	ajout_produit	sirop de grenadine	
41	18/01/08	14:54:26	demande_eval	3 61 411	
42	18/01/08	14:54:40	consult_FS	E124	
43	18/01/08	14:54:58	demande_eval	61 411	
44	18/01/08	14:55:25	consult_cours	dosages - onglet_cours	
45	18/01/08	14:57:30	consult_cours	dosages - onglet_cours	
46	18/01/08	14:57:52	consult_cours	dosages - onglet_cours	
47	18/01/08	14:58:15	ajout_action	reference(23) - materiel = ; solution = eau ; commentaire = ; étape = 3	

Quantitatives, Numériques , Continues	Ordinales	Qualitatives, nominales, catégorielle,	Textuelles
Relatives à une mesure étalon Sur lesquelles des opérateurs arithmétiques peuvent être appliqués (ex : moyenne des âges)	Valeurs qui peuvent être ordonnées les unes par rapport aux autres Des opérateurs ne peuvent pas être appliqués sur ces variables	Valeurs qui ne peuvent pas être ordonnées les unes par rapport aux autres Des opérateurs ne peuvent pas être appliqués sur ces variables	Texte
Age Durée Note, Score Nombre d'actions	Classe d'âge, Rang des élèves, ...	Sexe Matière étudiée Niveau de rétroaction ...	Réponses à des questions ouvertes
Moyenne, médiane, min, max, quartile, corrélation, etc...	Fréquences, pourcentages , corrélation des rangs, médiane, quartile , moyenne b	Fréquences, pourcentages,	Analyse de contenu , lexicale, sémantique

Rôles dans l'analyse

Explicatives,
Facteurs expérimentaux
« Indépendantes »



A expliquer
« Dépendantes »

Variables qui peuvent avoir une influence sur le comportement d'une autre variable
Variables qui seront identifiées a priori comme un élément explicatif potentiel



Variables dont le comportement est influencé par une (des) autres variables

le « parcours »

le « groupe expérimental vs le groupe contrôle »



Le « score obtenu »

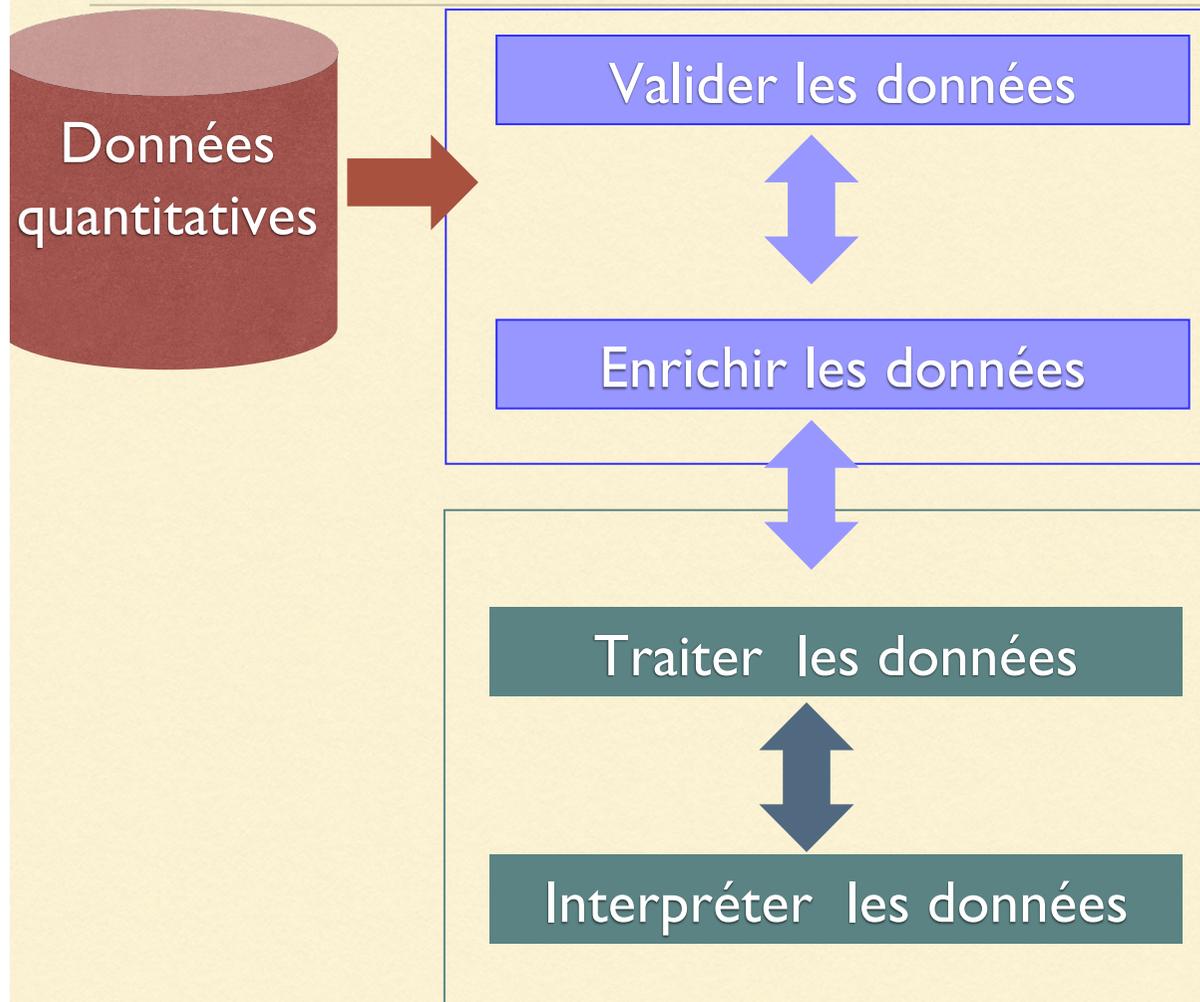
le « temps passé à consulter les cours »

La note d' « utilisabilité »

ANALYSE QUANTITATIVE

- Que pensez vous de cette formule ?
- Données + Algorithme de traitement = Résultats

ANALYSE QUANTITATIVE



Pré-traitement = 80%

- **Valider** identifier les outliers, les manquants, ...
- **Enrichir** les données créer de nouvelles variables
- **Traiter** : appliquer des méthodes statistiques pour donner du sens aux données recueillies
- **Interpréter** les résultats par rapport aux questions de recherche

ANALYSE QUANTITATIVE

BONNES PRATIQUES DE VALIDATION

- Mettre en relation les données mesurées avec les valeurs attendues, *Exemple : note des élèves comprises entre 0 et 20 ou entre 0 et 10.*
 - Si les valeurs mesurées sont quantitatives, rechercher les valeurs minimales et maximales
 - Identifier les individus pour lesquelles les valeurs sont manquantes
 - Croiser des variables entre elles, pour voir la cohérence, *Exemple : âge et date de naissance ; âge de l'élève et niveau scolaire*
- ⇒ **ne pas supprimer les données : garder le fichier de départ et créer un nouveau fichier de travail**
- Sur certains logiciels, les déclarer en données manquantes ,
 - Les corriger, en gardant une trace des corrections (pgm informatique, note sur word,) ,
 - Faire des traitements stats en filtrant les données

ANALYSE QUANTITATIVE

Validité ?

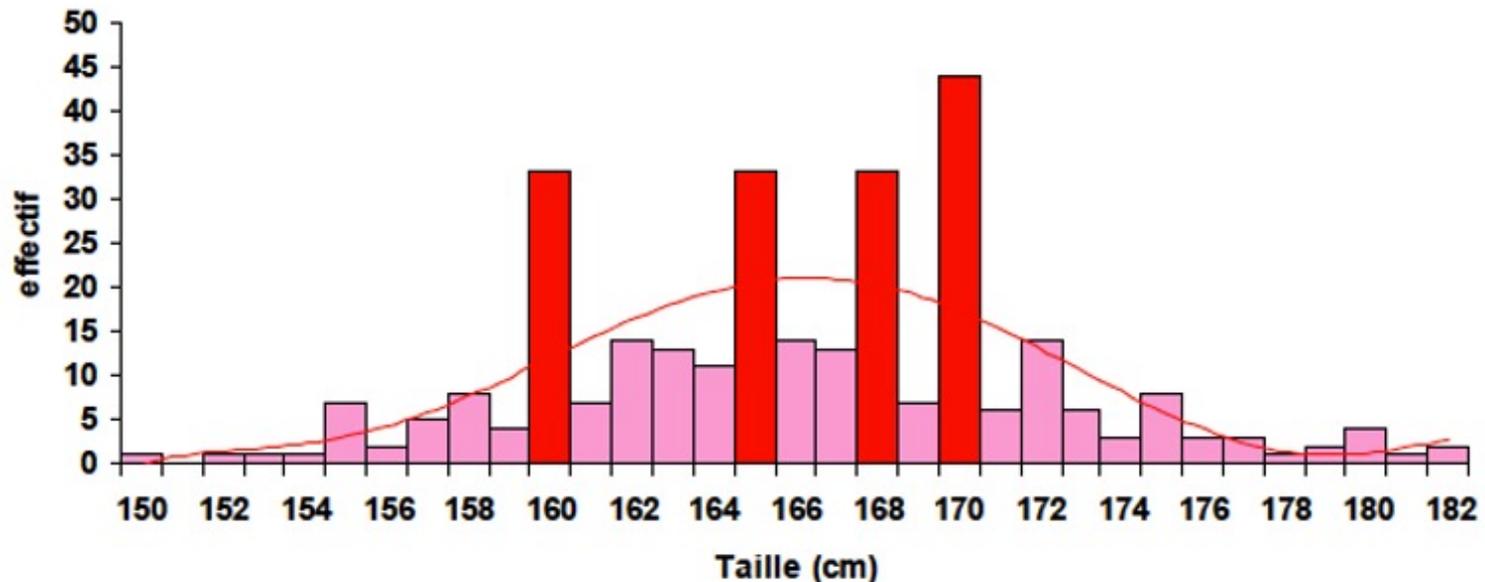
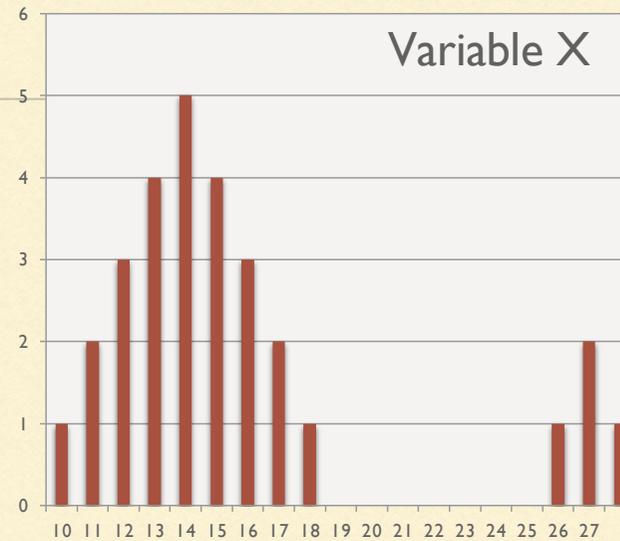


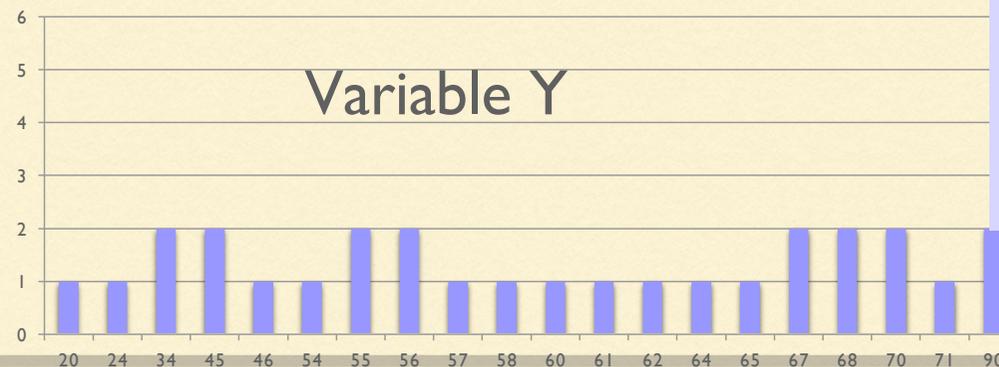
Figure 3.4. tailles auto-déclarées de 305 étudiantes de maîtrise. On observe des "effets de seuils" nets, avec une fréquence anormalement élevées de déclarations pour certaines tailles.

HISTOGRAMME DE DISTRIBUTION

valeurs mesurées	effectifs
10	1
11	2
12	3
13	4
14	5
15	4
16	3
17	2
18	1
19	0
20	0
21	0
22	0
23	0
24	0
25	0
26	1
27	2
28	1



valeurs mesurées Y	Effectifs
20	1
24	1
34	2
45	2
46	1
54	1
55	2
56	2
57	1
58	1
60	1
61	1
62	1
64	1
65	1
67	2
68	2
70	2
71	1
90	2



ANALYSE QUANTITATIVE

- INDICATEURS STATISTIQUES
- Lesquels connaissez vous ?

STATISTIQUES DESCRIPTIVES

Moyenne :
Classe 1 : 10,4
Classe 2 : 10,4

STATISTIQUES DESCRIPTIVES

Moyenne :	Médiane
Classe 1 : 10,4	Classe 1 : 11
Classe 2 : 10,4	Classe 2 : 10

STATISTIQUES DESCRIPTIVES

Moyenne :	Médiane	Etendue
Classe 1 : 10,4	Classe 1 : 11	Classe 1 : 11
Classe 2 : 10,4	Classe 2 : 10	Classe 2 : 1

STATISTIQUES DESCRIPTIVES

Moyenne :	Médiane	Etendue	Ecart type
Classe 1 : 10,4	Classe 1 : 11	Classe 1 : 11	Classe 1 : 3,61
Classe 2 : 10,4	Classe 2 : 10	Classe 2 : 1	Classe 2 : 0,53

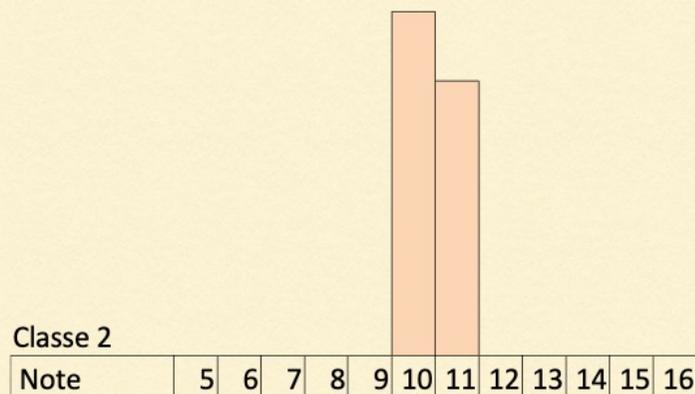
STATISTIQUES DESCRIPTIVES

Moyenne :
 Classe 1 : 10,4
 Classe 2 : 10,4

Médiane
 Classe 1 : 11
 Classe 2 : 10

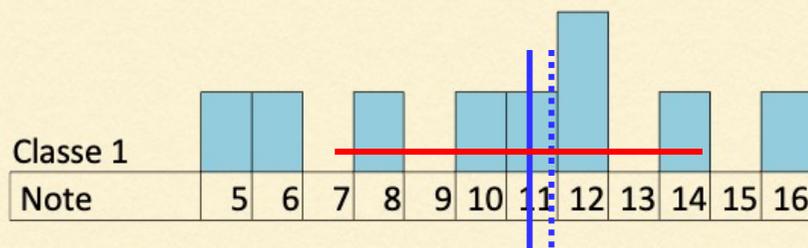
Etendue
 Classe 1 : 11
 Classe 2 : 1

Ecart type
 Classe 1 : 3,61
 Classe 2 : 0,53



classe 1	classe 2
8	10
12	11
16	10
5	10
6	11
12	11
14	11
11	10
10	10

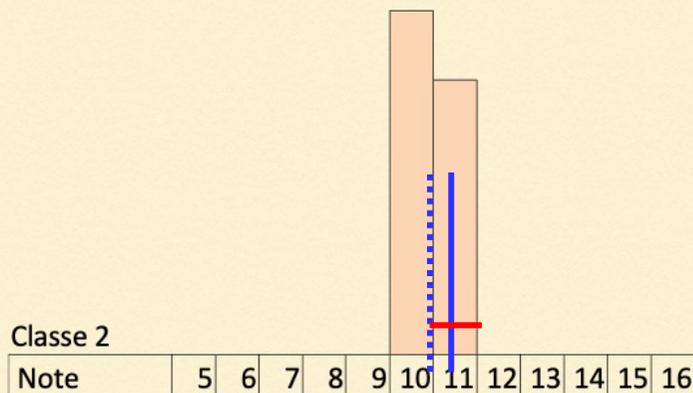
STATISTIQUES DESCRIPTIVES



Positionnement

Moyenne :
Classe 1 : 10,4
Classe 2 : 10,4

Médiane
Classe 1 : 11
Classe 2 : 10



Dispersion

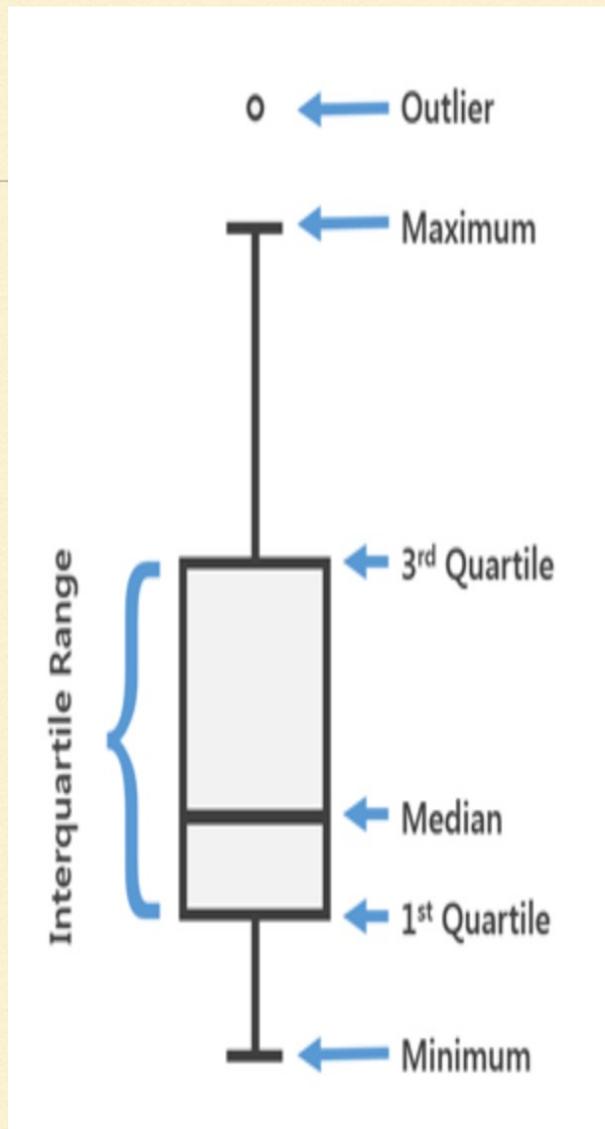
Etendue
Classe 1 : 11
Classe 2 : 1

Ecart type
Classe 1 : 3,61
Classe 2 : 0,53

FREQUENCES, POURCENTAGES ET QUARTILES

- Fréquences : nombre d'occurrences d'une valeur
- Pourcentages : nombres d'occurrences d'une valeur/effectif total *100
 - Valide si effectif total (n) ≥ 100
- Quartiles : découpage de la distribution en 4 parties égales
 - 1er quartile est la donnée de la série qui sépare les 25 % inférieurs des données (Q1) ;
 - 2e quartile est la donnée de la série qui sépare les 50 % inférieurs des données (Q2 \leftrightarrow Médiane) ;
 - 3e quartile est la donnée de la série qui sépare les 75 % inférieurs des données (Q3) ;
 - le 4e quartile est le maximum (Q4)
- La distance interquartile = $Q3 - Q1$
- Déciles : découpage de la distribution en 10 parties égales
- Percentiles : découpage de la distribution en 100 parties égales

BOX PLOT



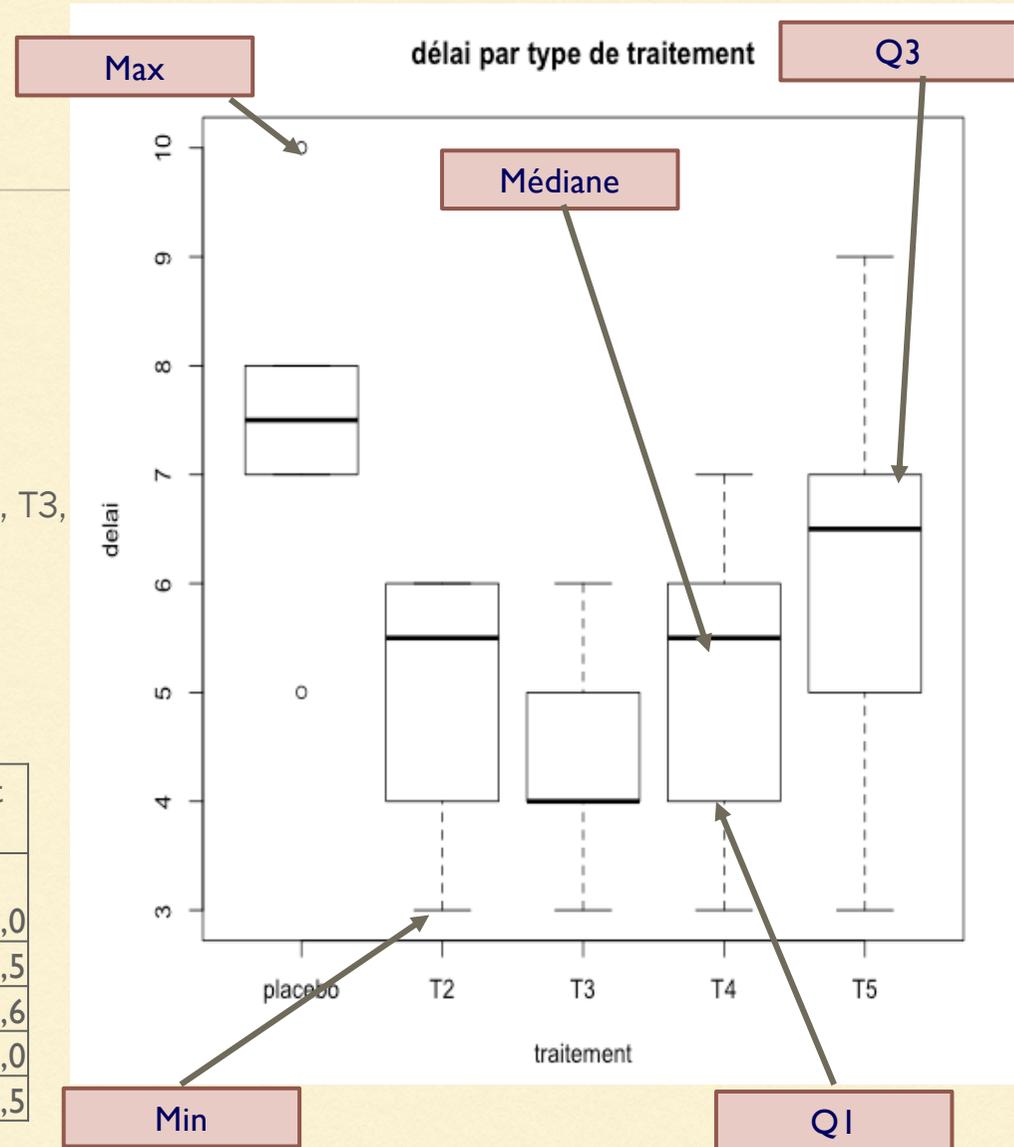
-
- Barre du bas : $Q1 - 1,5 \text{ IQ}$
- Deuxième barre : $Q1$
- Troisième barre : médiane
- Quatrième barre : $Q3$
- O : à plus de 1,5 distance inter quartile (IQ)
- * : à plus de 3 distances inter quartile

BOX PLOT

■ Délai de guérison et type de traitements

- Délai de guérison en jours : Quantitative
- Traitement : Qualitative à 5 modalités (placebo, T2, T3, T4, T5)

	Min	Q1	Median	Mean	Q3	max	Etendue	Ecart type
Placebo	5	7,0	7,5	7,5	8,0	10,0	5,0	2,0
T2	3	4,3	5,5	5,0	6,0	6,0	3,0	2,5
T3	3	4,0	4,0	4,3	4,8	6,0	3,0	2,6
T4	3	4,3	5,5	5,2	6,0	7,0	4,0	3,0
T5	3	5,3	6,5	6,2	7,0	9,0	6,0	3,5



DÉCILES ET PERCENTILES

Nombre de ressources ouvertes lors de l'exercice de math										
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Groupe 1	2	3	4	15	10	11	12	15	20	35
Groupe 2	2	7	10	15	20	21	22	24	28	29
Groupe 3	0	0	0	0	0	0	0	0	4	20

Qu'observez vous ?
Quelles différences entre les groupes ?
Que faire du groupe 3 ?
Comment recoder cette variable ?

TABLEAUX CROISÉS

- Mettre en relation deux variables qualitatives

Fréquences	Men	Women	Total
Humanities	40	100	140
Natural Science	110	100	210
Social sciences	80	140	220
Total	230	340	570
Pourcentages	Men	Women	Total
Humanities	7,0	17,5	24,6
Natural Science	19,3	17,5	36,8
Social sciences	14,0	24,6	38,6
Total	40,4	59,6	100,0

TABLEAUX CROISÉS

- Mettre en relation deux variables qualitatives

% lignes	Men	Women	Total
Humanities	28,6	71,4	100
Natural Sciences	52,4	47,6	100
Social sciences	36,4	63,6	100
Total	40,4	59,6	100
% colonnes	Men	Women	Total
Humanities	17,4	29,4	24,6
Natural Science	47,8	29,4	36,8
Social Sciences	34,8	41,2	38,6
Total	100,0	100,0	100

TESTS STATISTIQUES

- **Tester quoi ?**
 - Comparer des indicateurs statistiques
 - Moyennes, Proportions, Variance, ...
 - Tester des liaisons entre variables
 - Corrélacion, indépendance, ...
 - Tester des modèles
 - Régression (le poids est une fonction du poids)

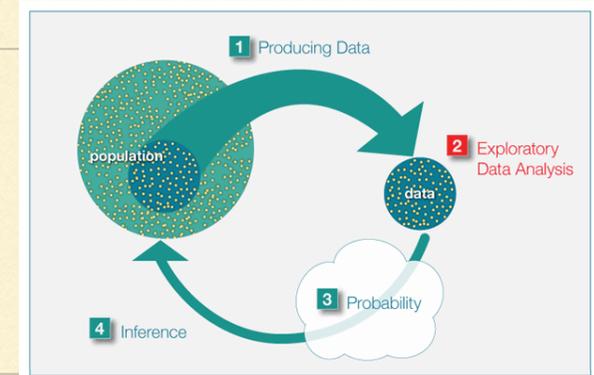
CHOIX DES TESTS

- Méthodes de tests sont liées
 - À la combinaison rôles et types des variables
 - Aux distributions des variables (normales et non normales)

QUELQUES TESTS

			Variable(s) explicative(s)
		Condition d'application	Qualitative
<i>Type de la variable à expliquer</i>	Qualitative		Test d'indépendance du Chi ² (couleur des yeux et couleur des cheveux)
	Quantitative	La variable à expliquer SUIT une loi normale (test de shapiro)	Test de Student (2 groupes) Anova (plusieurs groupes)
	Quantitative	La variable à expliquer NE SUIT PAS une loi normale (test de shapiro)	Test de Wilcoxon ou test de Kruskal et Wallis

LES TESTS STATISTIQUES



HYPOTHESES STATISTIQUES

H0 (hypothèse nulle) : moyenne des notes des garçons = moyenne des notes des filles.

H1 (hypothèse alternative) : moyenne des notes des garçons \neq **moyenne** des notes des filles.

=> Posent deux questions

Est ce que la moyenne des notes que j'ai observée dans la classe est différente ou non entre les garçons et le filles ?

Est ce que résultat est généralisables d'autres classes ? Voire à l'ensemble des individus ?

STATISTIQUE : l'étude de l'égalité de deux moyennes suit une distribution d'un t de Student.

A partir des données observées => calcul le t de student observé (To)

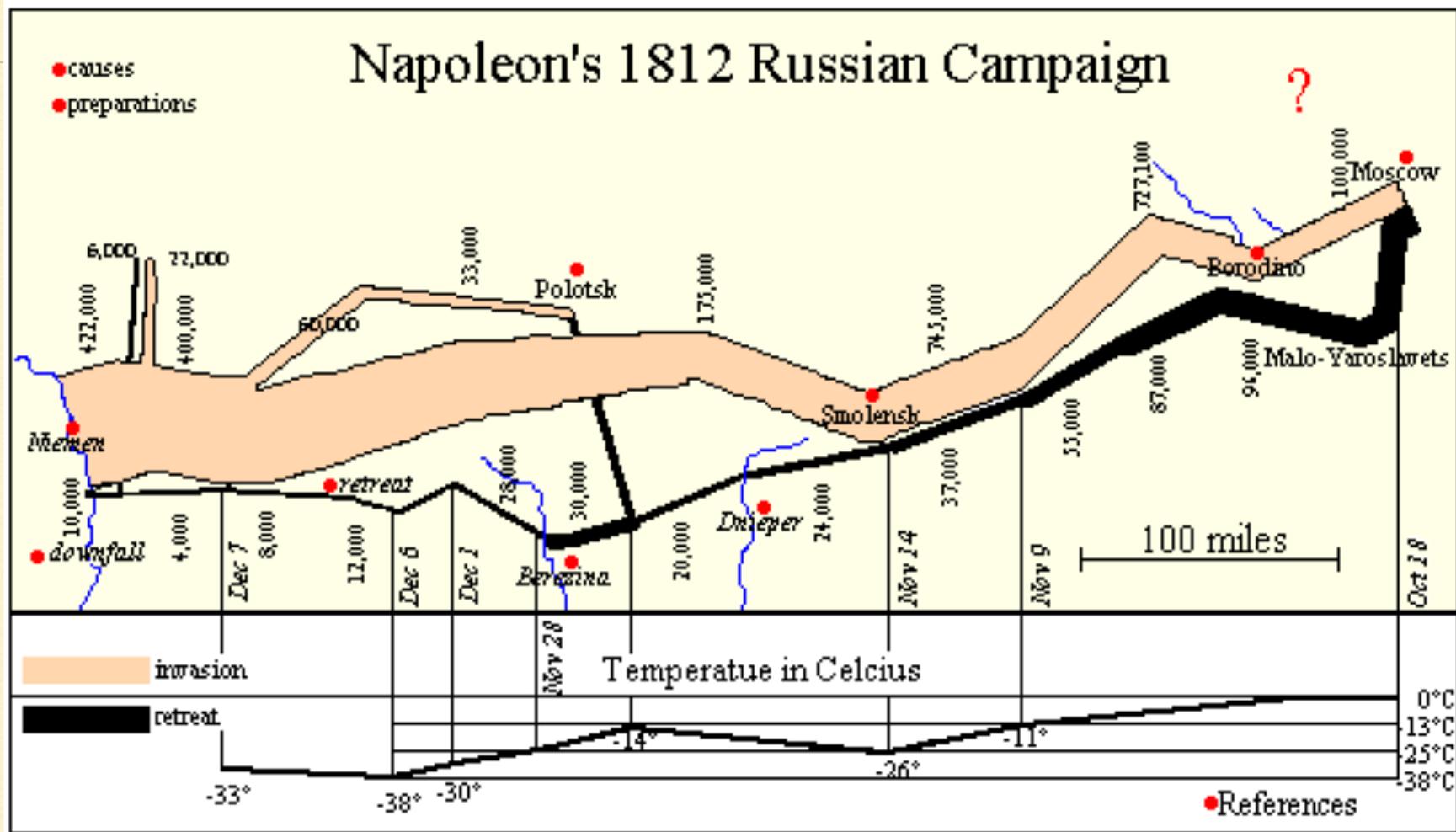
Compare ce To à un t théorique qui serait obtenu dans le cas de l'hypothèse H0.

Permet d'obtenir la probabilité d'obtenir un t de student supérieur à celui que j'ai observé dans ma classe , si on refaisait des relevés de notes dans une autre classe.

CONCLUSIONS

- = Si le t observé est supérieur au t théorique au seuil de 1% : le test est significatif au seuil de 1% Autrement dit, c'est la probabilité de trouver un t Observé supérieur au t théorique dans le cas d'une autre classe
- = Le langage habituel : rejeter H0 : autrement dit les moyennes ne sont pas égales
- = Le plus simple à comprendre : on accepte H1 avec un risque de se tromper égal à 1%.

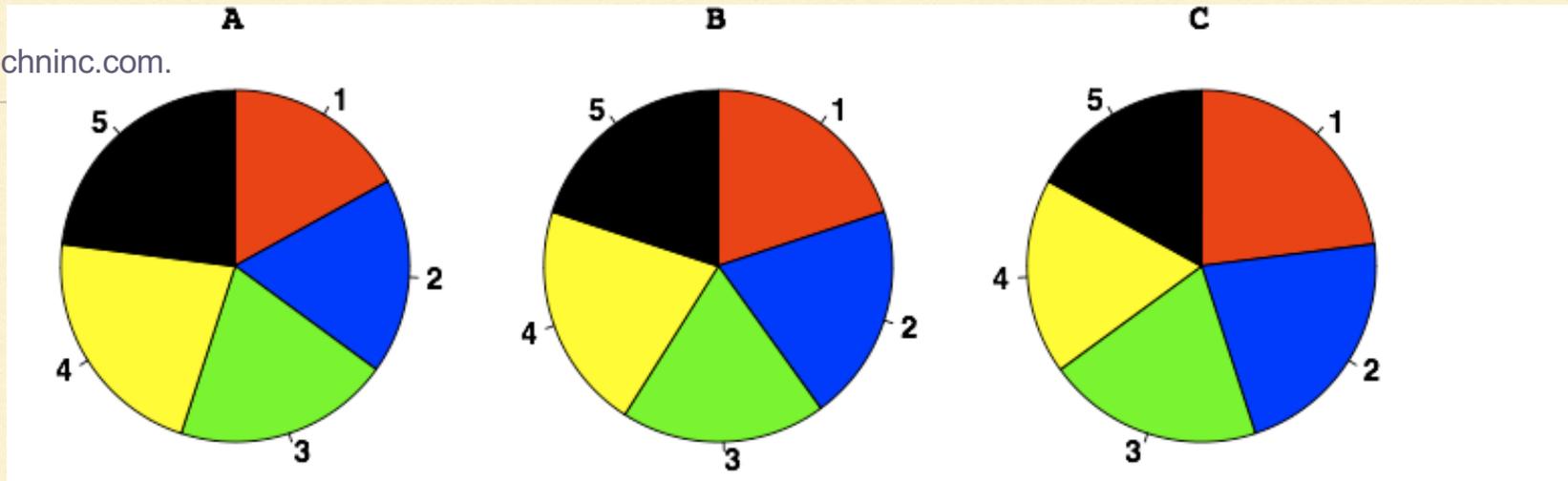
LES CARTES DE MINARD



VISUALISATION DES DONNÉES

« SAVE THE PIES FOR DESSERT », STEPHEN FEW 2007.

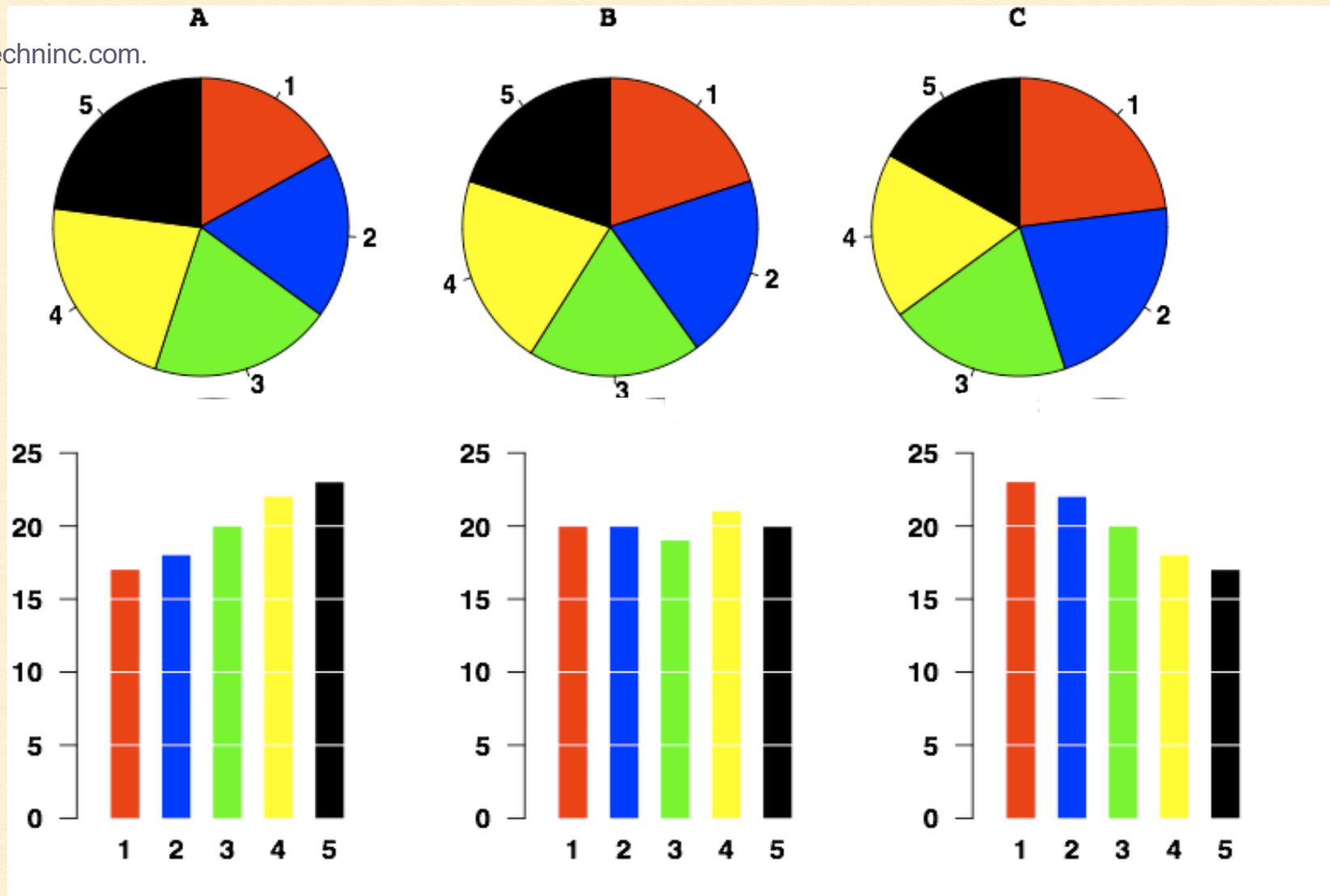
Source : innovatechninc.com.



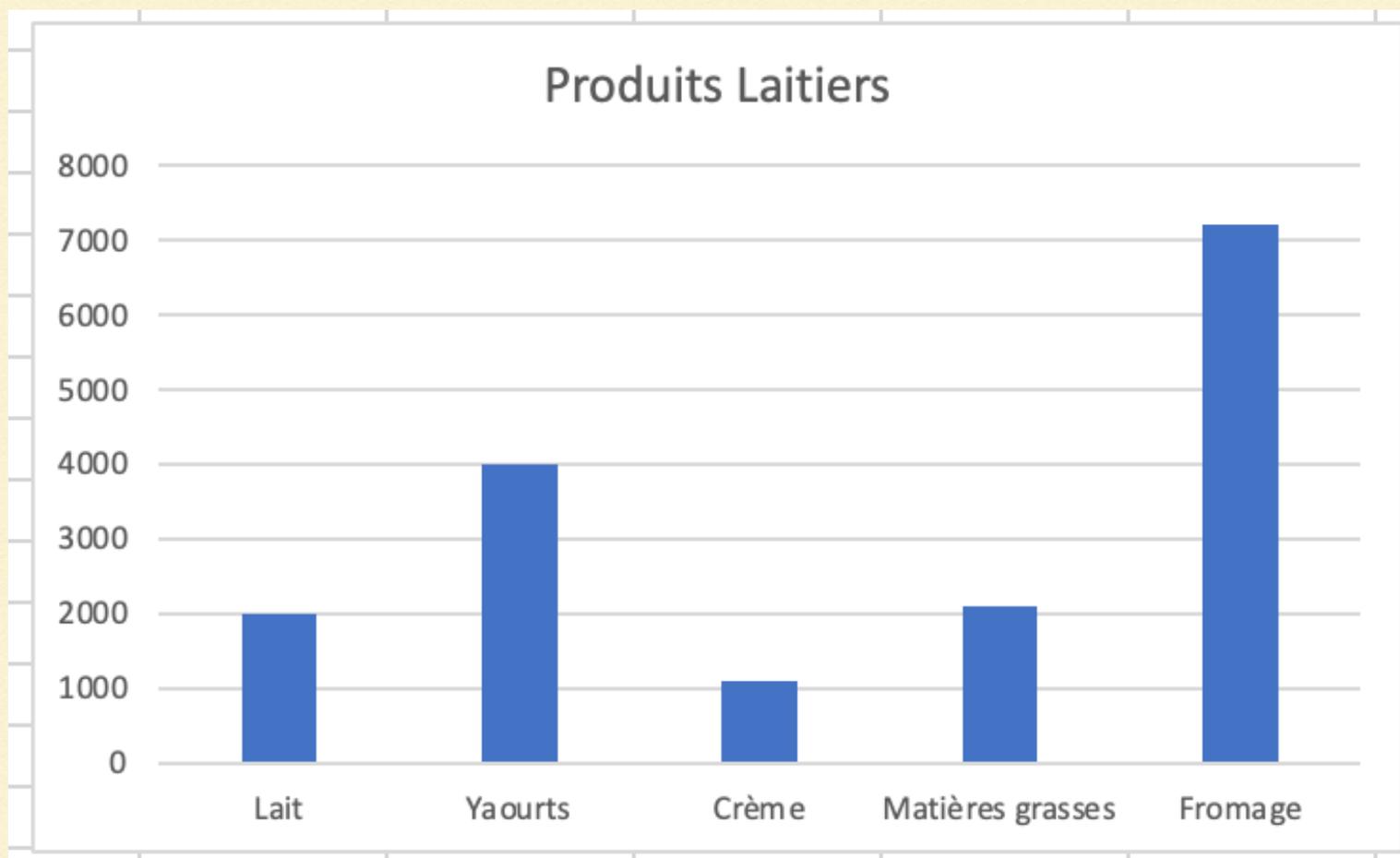
VISUALISATION DES DONNÉES

« SAVE THE PIES FOR DESSERT », STEPHEN FEW 2007.

Source : innovatechninc.com.



ELEMENTS POUR UN BON GRAPHIQUE



ELEMENTS POUR UN BON GRAPHIQUE

